



F0102593 RR (1)

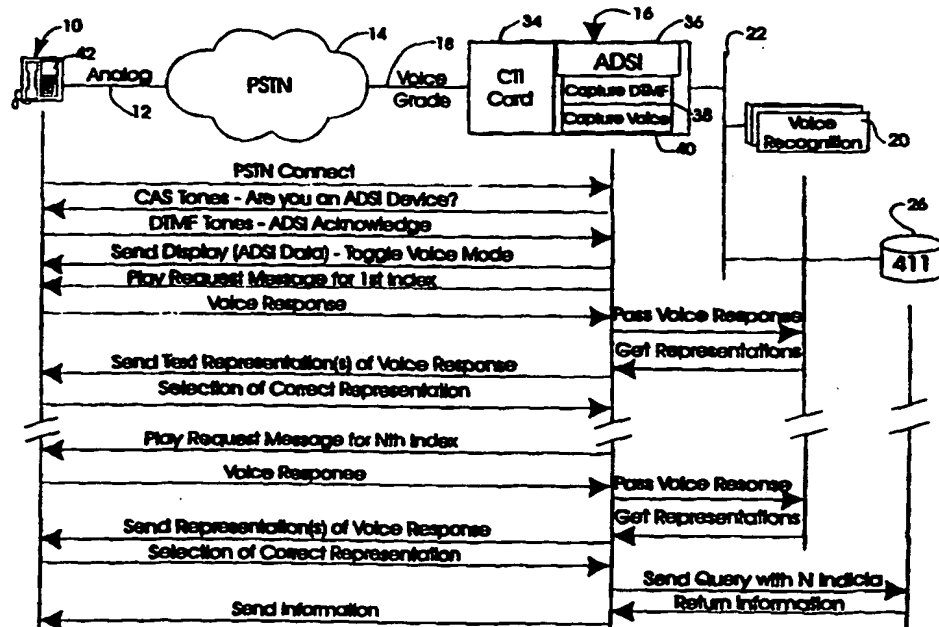
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G10L 5/06		A1	(11) International Publication Number: WO 99/21171
			(43) International Publication Date: 29 April 1999 (29.04.99)
(21) International Application Number: PCT/CA98/00978		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 21 October 1998 (21.10.98)		Published With international search report.	
(30) Priority Data: 2,219,008 21 October 1997 (21.10.97) CA			
(71) Applicant: BELL CANADA [CA/CA]; Bureau 1500, Côte du Beaver Hall, Montreal, Quebec H2Z 2S4 (CA).			
(72) Inventors: FORTIER, Stéphane; Avenue 15e #7760, St.-Georges, Quebec G5Y 5C2 (CA). REID, Colin, A.; Apartment 2205, Kent Street 199, Ottawa, Ontario K2P 2K8 (CA). WILLIAMS, L., Lloyd; Brodeur Crescent 15, Kanata, Ontario K2L 1Z2 (CA). BOUCHARD, Jean; Guyon 74, Beauport, Quebec G1B 3E7 (CA).			
(74) Agents: WOOD, Max, R. et al.; Swabey Ogilvy Renault, Suite 1600, 1981 McGill College Avenue, Montreal, Quebec H3A 2Y3 (CA).			

(54) Title: A METHOD AND APPARATUS FOR IMPROVING THE UTILITY OF SPEECH RECOGNITION

(57) Abstract

A method and apparatus for improving the utility of speech recognition is described. The method involves capturing a spoken word, passing the spoken word to a speech recognition algorithm (20), receiving at least one text representation of the spoken word from the speech recognition algorithm, and passing a text representation of the spoken word to a display telephone (10) to permit the speaker to select the correct representation of the spoken word. The apparatus includes an access server (16) that communicates with the display telephone, a speech recognition algorithm that responds to queries from the access server and one or more databases (26-34) that likewise respond to queries from the access server. The method and apparatus are particularly useful in automating such functions as telephone directory services using display telephones. The advantage is the ability to completely automate directory services for owners of display telephones and to significantly broaden the applications for speech recognition as a tool in information retrieval and transaction processing.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

- 1 -

A METHOD AND APPARATUS FOR IMPROVING THE UTILITY
OF SPEECH RECOGNITION

TECHNICAL FIELD

5 This invention relates generally to the field of speech recognition and, in particular, to an apparatus and method for improving the utility of speech recognition, especially in respect of automated processes for retrieving information from a database.

10

BACKGROUND OF THE INVENTION

 The desirability of retrieving information from databases using speech recognition as a means for converting spoken words into indicia useful in retrieving
15 information has long been recognized. A principal application for this technology has been the partial automation of telephone directory assistance services. Telephone companies and telephone equipment manufacturers have invested considerable resources in developing
20 systems to reduce the labour costs associated with providing directory assistance services. Much of that investment has been in speech recognition algorithms designed to facilitate directory look-ups.

 Although speech recognition algorithms have
25 been consistently improved, they have to date failed to provide complete automation solutions for information retrieval applications such as directory assistance. Due to the nature of spoken language, speech recognition is inherently limited in its ability to discriminate between
30 words which are pronounced alike but spelled differently. The utility of speech recognition is further challenged

- 2 -

by the current mobility of the world population which contributes to a diverse ethnic mix and consequently a variety of accents and inflections in most urban centres. Consequently, most speech recognition algorithms, although finely tuned and inventively designed, are incapable of enabling complete automation of an information retrieval system.

The inherent limitations in speech recognition are readily understood. For example, humans as well as machines have difficulty in distinguishing between the sounds uttered for the letters B, C, D, E, G, P, T, V and Z. In addition, it is substantially impossible to determine the spelling of certain words, including names, based on their pronunciation. For example, the names John, Jon and Jean may all be pronounced similarly enough that discrimination of the true spelling is impossible. Likewise, the names Mary Ann, Maryanne and Marianne are simply impossible to differentiate as spoken words. Other examples too numerous to mention may be readily recited by speech scientists.

A need therefore exists for a method of improving the utility of speech recognition in order to permit the automation of functions which are usefully implemented using speech recognition technology.

SUMMARY OF THE INVENTION

It is an object of the invention to provide a method and apparatus for improving the utility of speech recognition to permit the automation of information retrieval systems which use speech recognition as the primary engine for information retrieval.

- 3 -

It is a further object of the invention to provide a method of implementing speech recognition that enables the automation of transactions that may be accomplished over the switched telephone network.

5 It is another object of the invention to provide a method and apparatus for improving the utility of speech recognition which is relatively easy to design and inexpensive to implement.

10 It is yet a further object of the invention to provide a method and apparatus of improving the utility of speech recognition that is designed to enable the complete automation of telephone directory assistance services.

15 In accordance with a first aspect of the invention there is provided a method of improving the utility of speech recognition of words spoken by a speaker in which a word spoken by the speaker is captured in electronic form and passed to a speech recognition algorithm (20) and at least one representation of the word is returned by the speech recognition algorithm
20 C H A R A C T E R I Z E D in that:

the at least one representation of the word is displayed for the speaker as text to permit the speaker to select a correct representation of the word from among
25 the at least one displayed representation.

In accordance with the second aspect of the invention, there is provided apparatus for improving the utility of speech recognition of words spoken by a speaker, comprising a computer (16) enabled to receive
30 voice and data signals over a communications link (18), the computer being programmed to prompt the speaker for

- 4 -

spoken words which are received from the communications link as voice signals and to pass the spoken words to a speech recognition algorithm (20) which returns representations of the spoken words to the computer

5 C H A R A C T E R I Z E D in that:

the computer is further enabled to pass the representations of the spoken words to a voice terminal with a display surface which displays the representations for the speaker to permit the speaker to select a correct
10 representation of the spoken words.

The invention therefore provides a method and an apparatus for improving the utility of speech recognition and enables a much broader application of speech recognition technology, especially in the
15 implementation of services which entail the retrieval of information from databases. In accordance with the method, a database query is assembled by prompting a user to verbally indicate each of several names which may be used as indexes for retrieving a record of interest from
20 a database. Each name may consist of one or more spoken words. The names are preferably requested in sequence and each name is preferably verified by passing the spoken words to a speech recognition algorithm which returns at least one text representation of the spoken
25 name. The representations of the spoken name are then presented to the speaker who is permitted to select the correct representation of the spoken name. After all of the names required for a query have been correctly identified, a query is assembled and submitted to the
30 database. This permits an accuracy of information

- 5 -

retrieval which was heretofore unattainable using speech recognition alone.

The apparatus in accordance with the invention consists of voice terminals having display surfaces for displaying characters and a computer which may be accessed by the voice terminals. The computer in turn has access to a speech recognition algorithm and a database which stores the information of interest. Software enables the computer to prompt the user to utter the names required as indicia for locating a record of interest in the database. Software also enables the computer to submit captured voice signals to the speech recognition algorithm which returns one or more textual representations of the spoken name. Graphical representations of the spoken name are displayed as text on the display surface of the voice terminal to permit the user to select the correct representation. The invention may therefore be inexpensively implemented to enable a wide variety of applications.

The method and apparatus in accordance with the invention is particularly adapted to providing completely automated directory services to individuals having display telephones. The display telephones are preferably adapted to conform to the Analog Display Services Interface (ADSI) standard FR-12 developed by Bellcore. The computer is preferably a server which may be accessed by a dial-up voice-grade connection. The speech recognition algorithm may reside on the same server or may reside on another server in a local or wide-area network. Preferably, at least one speech recognition algorithm is provided in every region of a

- 6 -

telephone network in order to permit regional training for the recognition of locality names as spoken by local speakers. The directory database, on the other hand, is preferably accessed through a wide-area network and
5 centralized to eliminate maintenance duplication and maximize accuracy.

It will be well understood by those skilled in the art that this technology may be used in many other applications where information is usefully retrieved or
10 transactions are conducted using spoken language. The method and apparatus in accordance with the invention may therefore also be used, for example, to implement a voice order system for telephone retail sales operations, an automated voice reservation system for hotel
15 accommodations, and many other applications too numerous to mention.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be further explained by
20 way of example only and with reference to the following drawings wherein:

Fig. 1 is a schematic diagram of a portion of a network which includes the apparatus in accordance with the invention arranged in a configuration suitable for
25 retrieving information from telephone directory databases;

Fig. 2 is a schematic diagram of a portion of a display telephone which may be used to implement the method in accordance with the invention;

30 Fig. 3 is a schematic diagram showing the information flow during the retrieval of a record from a

- 7 -

database using the method and apparatus in accordance with the invention;

Figs. 4a-e are flow charts of one way in which the logic executed by the apparatus in accordance with the invention may be organized; and

Fig. 5, which appears on sheet 2 of the drawings, is a block diagram of a telephone directory database accessed using the method and apparatus in accordance with the invention.

10

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

This invention relates to a method and apparatus for improving the utility of speech recognition using text prompts to verify spoken words converted to text by a speech recognition algorithm. The method and apparatus are particularly well adapted for providing automated directory services in the telephone network. The invention may also be used for the provision of more complex automated services including automated speech ordering for product sales, automated system for the reservation of accommodations, and the like.

Fig. 1 shows a schematic diagram of a portion of a network which includes the apparatus in accordance with the invention. The apparatus is configured to provide directory services to subscribers having a display telephone 10. The display telephone 10 is preferably a telephone which conforms to the Analog Display Services Interface (ADSI) standard. Such display telephones are available for example from Nortel of Ottawa, Ontario, Canada; Phillips Home Services Inc. of Burlington, MA, USA; or Cidco Inc. of Morgan Hill, CA,

30

- 8 -

USA. The ADSI display telephone 10 is connected by an Analog line 12 to the Public Switched Telephone Network (PSTN) 14. It should be understood by those skilled in the art that although the invention described below is described with reference to an ADSI display telephone 10, the invention may also be implemented using a digital telephone or any other voice terminal which has a character display capability and responds to commands received from a remote computer in the client/server model.

The apparatus in accordance with the invention includes one or more access servers 16 connected to the PSTN 14 by voice grade connections 18 which may be lines or trunks, for example ISDN PRI trunks, which are well known in the art. The access servers 16 may query speech recognition algorithms 20 in order to convert words spoken by subscribers using the ADSI display telephone 10 into one or more representations of the spoken words in a manner well known in the art. The speech recognition algorithms 20 may be resident on the access server 16 or resident on other servers connected to a local or a wide area network 22. The speech recognition algorithm 20 may be for example an algorithm such as the one described in United States patent No. 5,479,488 which issued to the applicant on December 26, 1995 or one such as described in United States patent No. 5,644,680 which issued to Northern Telecom Limited on July 1, 1997.

The local area network 22 is connected to a wide area network 24 in a manner well known in the art. The wide area network 24 provides access to a plurality of directory databases which include, for example, a

- 9 -

"411" database of residential numbers 26, a Yellow Pages database 28, a database of business numbers 30, a database of 800 numbers 32, and a global database 34 which may include a variety of information respecting entities for which records exist. The databases 26-34 are exemplary only of the types of databases from which information may be retrieved using the improved speech recognition method and apparatus in accordance with the invention.

Fig. 2 shows a portion of an ADSI display telephone 10 which constitutes a part of the apparatus useful in practising the methods in accordance with the invention. The displayed portion of the ADSI display telephone 10 includes a display surface 42 which is enabled to display text characters and simple graphics. Located on each lower side of the display surface 42 are "soft keys" 44a-f. The function of each soft key 44a-f is allocated dynamically with each display on the display surface 42. In this example, soft key 44a enables the user to select one of the four options displayed. If soft key 44a is pressed with the display surface 42 in the condition illustrated, the subscriber will select the first name "Steven". If the subscriber selects soft key 44b, a retry signal will be sent to the access server 16 (see Fig. 1) and the process of selecting a first name will be reinitiated, as will be explained below in more detail with reference to Figs. 4a-4e. If the subscriber presses soft key 44d, the current search will be abandoned and the subscriber will be provided with options for disconnecting from the directory service or commencing a new search, for example. A cursor 46

- 10 -

indicates an item to be selected by the soft key 44a. The position of the cursor 46 on the display surface 42 may be changed using the cursor control keys 48 in a manner well known in the art.

5 Six function keys 50 are generally provided on the ADSI display telephone 10. The functions of those keys are manufacturer dependent and are not relevant to the invention.

Fig. 3 is a schematic diagram showing an
10 information flow during the retrieval of a record from a database using the method and apparatus in accordance with the invention. As explained above with reference to Fig. 1, an ADSI display telephone 10 is connected by an Analog line 12 to the PSTN 14. The access server 16 is
15 connected by a voice grade connection 18 to the PSTN 14. In order to connect to the access server 16, a subscriber dials a predetermined number using the ADSI display telephone 10. The access server 16 includes a computer telephony integration (CTI) card 34 that complies with
20 the Bellcore FR-12 standard available, for example, from the Dialogic Corporation of Parsippany, NJ, USA. The access server 16 also includes a processor and other hardware well known in the art which are standard components of a computer suitable to function as a server
25 in a local or a wide area network. The access server 16 is equipped with applications to implement an ADSI interface to handle communications with the ADSI display telephone 10 in a manner in accordance with the Bellcore Standard FR-12 Analog Display Services Interface. The
30 ADSI application 36 includes a Dual Tone Modulated Frequency (DTMF) capture section 38 and a voice capture

- 11 -

section 40. Each of the DTMF capture section 38 and the voice capture section 40 monitor the voice grade line 18 to detect any DTMF or voice signals received on the voice grade line 18. The DTMF capture section 38 has
5 precedence and extracts any DTMF tones generated by the ADSI display telephone 10, in a manner well known in the art.

Fig. 3 illustrates a simplified rendition of the information exchanged between the ADSI display
10 telephone 10, the access server 16, the speech recognition algorithm 20, and the 411 residential number directory database 26 when a subscriber uses the ADSI display telephone 10 to obtain the telephone number of a person from the 411 directory. As explained above, in
15 order to obtain a telephone number from the 411 directory 26, the subscriber dials a pre-assigned number which connects the ADSI display telephone 10 to the access server 16. When the access server 16 receives a call over the voice grade connection 18, it returns CAS
20 tones in accordance with the ADSI standard to interrogate the calling voice terminal to ensure that it is an ADSI device. If the telephone is an ADSI device such as ADSI display telephone 10, it responds to the CAS tones and returns DTMF tones in an ADSI acknowledgement message
25 which enables the access server 16 to continue the transaction. If no DTMF tones in accordance with the ADSI standard are received from the connected telephone, the access server 16 plays back an appropriate service denial message (see Fig. 4a) and may announce alternate
30 directory service options to the subscriber.

- 12 -

Assuming that the required DTMF tones for an ADSI display acknowledgement are received from the ADSI display telephone 10, the access server 16 sends display data in ADSI data format to the ADSI display telephone 10. The display data is accepted by the ADSI display telephone 10 and displayed on the display area 42 of the telephone to provide the subscriber with a visual introduction to the service. The ADSI display data is followed by a voice mode toggle in accordance with the ADSI standard to switch the ADSI display telephone 10 from data to voice mode in a manner established by the ADSI standard. The access server 36 subsequently plays a first request message for voice input from the subscriber. In the abbreviated information exchange illustrated in Fig. 3, the access server 16 immediately prompts the subscriber for a first index for the directory lookup. In real implementations, the first request message may be, for example, a request for a preferred language. That request may be followed by a request for the name of the directory of interest, for example any one of the databases 26-34 shown in Fig. 1. For the sake of clarity of illustration, Fig. 3 does not show those initial setup messages.

The request for the first index will, for example, be a request for the state or province in which the person resides for whom a telephone number is sought. The subscriber may therefore respond with a spoken name such as "Ohio". The subscriber's voice response is captured by the voice capture section 40 of the access server 16 and passed to the speech recognition algorithm 20 which processes the voice signal in a manner

- 13 -

well known in the art to produce one or more representations of the voice signal. The speech recognition algorithm 20 returns one or more representations of the voice signal to the access server 16. The access server 16 sends a text representation of the voice response to the ADSI display telephone 10 which displays the text representations on the display surface 42. As explained above with reference with Fig. 2, the ADSI display telephone 10 includes a plurality of soft keys 44a-f and cursor control keys 48 which permit the subscriber to select the correct text representation of the voice response or to restart the process. In this instance, the speech recognition unit 20 may have returned "Ohio" and "Iowa". Since the subscriber is interested in the state of Ohio, the subscriber operates the soft keys to select "Ohio". The selection is returned as DTMF tones to the access server 16. The DTMF capture section 38 of the access server 16 captures the DTMF tones and determines that the first index to be used for a database query is the state of "Ohio".

The first index is stored and the process is repeated for the remainder of the indexes required to retrieve a record from the database. In the instance of a residential telephone directory, the indexes generally further include city, last name and first name or initial. In the example shown in Fig. 2, the "Nth" index is first name. The first name of the party being sought may be for example "Steven". After the access server 16 plays a request for the first name of the party, the user responds with "Steven". The access server 16 captures

- 14 -

the voice signal and forwards it to the speech recognition unit 20 which returns four responses, namely: "Stephan", "Steven", "Stevan", and "Stevon". Using soft keys 44a-f and cursor control keys 48 on the display telephone 10, the user selects "Steven" (see Fig. 2) and that selection is returned to the access server 16.

The access server 16 now is in possession of all the indicia required to formulate a query to the database 26. The access server 16 therefore formulates a query which includes the "N" indicia and forwards the query to the database 26 over the Wide Area Network 24. On receipt of the query, the 411 directory of residential numbers 26 retrieves one or more records matching the "N" indicia. The one or more records are returned in an information package to the gateway server 16 which sends the information package to the ADSI display telephone 10. On receipt of the information package, the one or more records are displayed on the display surface 42 to permit the user to select the correct record using name and address as selection criteria. Thus, a completely automated directory service may be provided.

It will be understood by those skilled in the art that the speech recognition algorithm 20 may not always be able to return a correct representation of a spoken word. This may occur if the subscriber has an unusual accent or a search is conducted for a very unusual name which the speech recognition algorithm is not equipped to recognize. It is therefore preferable that another way of supplying a name be provided. This may be accomplished in several ways. After an unsuccessful attempt at providing a correct

- 15 -

representation of a name, one of the soft keys 44a-f may be enabled to permit the subscriber to verbally spell the name, in which case a speech recognition algorithm 20 must be provided which has an alpha recognition capability to permit it to recognize the letters of the alphabet in the language of interest. Normally, a separate alpha recognition algorithm is used for this purpose. Alternatively, one of the soft keys 44a-f may be enabled to permit the subscriber to manually spell the name using the dialing key pad of the ADSI display telephone 10. Manual spelling may be accomplished, for example, by programming the ADSI display telephone 10 and the access server 16 to recognize different letters of the alphabet by the number of times in succession that a dialing key is pressed. For example, the "2" key pressed once equals "A" and pressed twice (2,2) equals "B", and the "9" key pressed three times (9,9,9) equals "Y". "Q" and "Z" must be separately accommodated. They may, for example, be assigned to the "7" and "9" keys respectively, in which case "7" pressed once equals "P", pressed twice equals "Q" (7,7), pressed three times equals "R" (7,7,7), and pressed four times equals "S" (7,7,7,7). The same sequence is used for the "9" key and the letters "W", "X", "Y" and "Z", respectively. If the apparatus in accordance with the invention is thus enabled to accept input in another way besides spoken words, a service such as a directory assistance service may be completely automated and still provide an acceptable level of user satisfaction. In real applications, of course, one of the soft keys 44a-f would normally be programmed to provide optional access to a

- 16 -

live operator if a predefined number of restarts were encountered during a directory search.

In order to further illustrate the best mode of practising the invention, Figs. 4a-e show a flow chart of a preferred arrangement of the logic enabled on the access servers 16 to permit retrieval of information from a database as shown in Fig. 3.

As shown in Fig. 4a, when an access server 16 is accessed using the voice grade connection 18, it verifies that the accessing telephone is ADSI compliant by sending CAS tones to the calling telephone in a step 60, as described above with reference to Fig. 3. In step 62, the access server 16 determines whether DTMF tones constituting an ADSI confirmation message are received from the telephone to which the CAS tones were sent. If ADSI confirmation tones are not received, the access server 16 plays a pre-recorded "invalid telephone" message in step 64 and drops the connection in step 66, which ends processing. The invalid telephone message played in step 64 may include an announcement of alternate services which may be accessed by the caller without an ADSI compliant telephone.

If the calling telephone is determined to be ADSI compliant in step 62, an initialization process begins which may include, for example, a request for a language of preference in step 68. In step 70, the access server 16 evaluates the caller's verbal response to determine whether it is a valid option respecting choice of language. If it does not match one of the valid options, a retry limit is incremented and the limit is tested to determine whether it has exceeded a

- 17 -

predefined limit in step 72. If the predefined limit has not been exceeded, an announcement is played requesting that the user please try again to select one of the languages in which the service is offered in step 74. If the predefined limit is found to be exceeded in step 72, a pre-recorded announcement that the language choice is not supported is played in step 76 and processing ends. If a valid response is received in step 70, the caller is requested in step 78 to request a database selection. The database selection requested in step 78 may be a request for a voice response or a request for a soft key response using the soft keys 44a-f (see Fig. 2). In step 80, a determination is made respecting whether a valid database has been selected. If an invalid response is received, a retry limit is incremented and the limit is tested to see whether a predefined limit is exceeded in step 82. If the limit has not been exceeded, the caller is requested to try again in step 84. If the retry limit has been exceeded the caller is played a pre-recorded message indicating that an unknown database has been repeatedly requested in step 88 and the call is terminated. When a valid database is selected, an initialization process is undertaken in step 86 to set pointers to the speech recognition algorithm and the database required to supply the desired information to the caller using the ADSI display telephone 10.

As explained above with reference to Fig. 3, the process in accordance with the invention for retrieving information from a database involves the steps of determining a value for a predefined number of indicia used to locate records in the database and formatting a

- 18 -

query after the user has confirmed a correct value for each of the predefined indicia. Figs. 4b-e illustrate the preferred logic involved in that process.

As shown in Fig. 4b, the process begins by
5 initializing a speech recognition algorithm for a current index being sought from the caller in step 90. It is well understood in the art that in order to improve the efficiency of speech recognition, a predefined table of values are normally associated with each index so that
10 speech recognition is more rapidly and accurately accomplished. The initialization in step 90 sets appropriate pointers to enable the access server 16 to conduct the required queries. After initialization, the access server 16 plays a pre-recorded announcement to
15 request a spoken value for the current index, for example the state or province of residence of an individual for whom a telephone number is sought. In step 94, the access server 16 monitors the voice capture section 40 (see Fig. 3) to determine whether a response is received.
20 If no response is received within a predetermined time interval tested in step 96, a retry limit is incremented and a determination is made in step 98 whether the retry limit is exceeded. If not, a repeat of the request for a current index is undertaken in step 92. If the retry
25 limit is exceeded, an action such as a play session termination message in step 100 may be executed or the user may be provided with other options such as initiating a new query or selecting a new database.

Once a valid voice response to a request for
30 the current index is received in step 94, the voice response is sent to the speech recognition algorithm in

- 19 -

step 102 and the access server 16 waits for a representation of the voice response to be returned in step 104. If an empty response is returned, a pre-recorded announcement is played in step 106 to inform the caller that the speech recognition has failed and the caller is preferably presented with alternate entry options in step 108 to permit the index to be entered in another way, as will be explained below with reference to Fig. 4e.

Assuming that at least one representation was determined to be returned in step 104, if only one representation was returned, it is sent to ADSI display telephone 10 in step 112 and displayed for the caller. In step 114, the caller is requested to verify that the representation is an accurate representation of the index displayed in step 112. The request for verification may be a pre-recorded message or a question displayed on the ADSI display telephone 10, or both. In step 116, the access server 16 determines whether the voice request is verified as correct. If it is not verified as correct, the caller is preferably presented with alternate entry options in step 118 as will be explained below with reference to Fig. 4e. If the caller verifies that the representation is a correct representation of the spoken index, the program branches to connector "C" (see Fig. 4c) where the index is stored and the process is repeated for the next index in step 148, as will be explained below with reference to Fig. 4c.

If it is determined in step 110 that more than one representation was returned from the speech recognition algorithm 20, the number of representations

- 20 -

is compared to a predefined limit in step 120. If it is determined that the number of representations exceeds the predefined limit, the recognition is considered to have failed and a pre-recorded speech recognition failed message is played in step 122 and the caller is presented with alternate entry options in step 124, which will be explained below with reference to Fig. 4e. If the number of representations does not exceed the predetermined limit, the first "N" representations are sent to the ADSI display telephone 10 in step 126 as well as an assignment of functions for the soft keys 44a-f (see Fig. 3). In step 128, a pre-recorded message requesting that the caller select a correct representation from the list is played. In step 130, the DTMF capture section 38 is monitored to determine whether a soft key has been pressed within a predetermined time. If the time has been determined to have expired in step 132, the announcement played in step 128 is repeated. This program loop may include a program cancel function (not illustrated) to ensure that excessive repeats are not executed. If a soft key is selected, a determination is made in step 134 as to whether the soft key is a request to display the next "N" representations. If so, a determination is made in step 136 as to whether the last "N" representations have already been displayed. If the last "N" representations have been displayed, a message is played in step 138 indicating that the last of the representations have been displayed. It will be understood by those skilled in the art that a backup function may also be provided to permit the user to redisplay the previous "N" selections, although this is

- 21 -

not illustrated. If all the representations have not been displayed, the next "N" selections are displayed in step 140 and the access server 16 returns to monitoring the DTMF capture section 38 to determine whether a soft key has been selected. Soft key monitoring includes determining whether the soft key indicates the selection of a representation in step 142. If it does not, a test is made to determine whether the soft key indicates a desire to exit in step 144. If it does not, an invalid soft key selected message is played in step 146 and soft key monitoring resumes. If the soft key is determined to indicate a desire to exit in step 144, the program returns to initializing the speech recognition algorithm for a current index at connector "A" (see Fig. 4b). If one of the representations is selected as a correct representation in step 142, the value of the representation is stored as an index in a predefined register in step 148 and the predefined registers are examined to determine whether the "Nth" index has been stored in step 150. If the "Nth" index has not be stored, the program is returned to connector "A" and the process of obtaining the next index is initiated. If the "Nth" index has been stored, the program proceeds to formulate a database query, the logic for which is shown in Fig. 4d.

After all indexes required for a database query have been accumulated and verified, the access server 16 formulates a database query in step 152 and sends the query to the appropriate database, database 26 for example (see Fig. 1). The access server 16 awaits a response from the database. The response may include

- 22 -

zero records or may include a large number of records depending on the indicia submitted in the query. If in step 154 the return is determined to be empty, the access server 16 may present an option to the caller in step 156 to broaden the query. Normally, this is accomplished by truncating the least significant index of the indicia or blanking the least significant index of the indicia. For example, suppose that a caller searches for the telephone number of John Doe in Toronto, Ontario and the query returns empty, the caller may be presented with an option to reformulate the query for "J. Doe" in Toronto, Ontario. The number, method and options for broadening queries are a matter of design choice and are well known in the art.

15 In step 158, the access server 26 monitors the connection for an indication as to whether the caller wishes the query to be broadened. If the caller does not wish the query to be broadened, the program preferably branches to connector "G" in which a pre-recorded message is played requesting that the user please select a new query or exit. If the caller requested that the query be broadened, the access server 16 formulates the new query in step 152 and sends it to the database. Thereafter step 154 is repeated. If an empty query is not returned, a determination is made in step 160 whether the number of returns exceeds a predetermined limit. If that limit is exceeded, a pre-recorded message is played in step 162 informing the caller that an excess number of returns have been received from the database and requesting whether the caller wishes those returns displayed on the ADSI display telephone 10. A response is monitored in

- 23 -

step 164 to determine whether a display of excess returns is requested. If so, a predetermined number of returns are sent to the ADSI display telephone 10 along with initialization of the soft keys appropriate for such display in step 166. Step 166 is also executed if a determination is made in step 160 that the number of returns does not exceed the predetermined limit. In step 168, the connection is monitored to determine whether a soft key indicates a "display more" command. If so, step 166 is repeated. If not, the caller has either located the record of interest or may initiate a new query. Therefore, in step 170, the connection is monitored by the access server 16 to determine whether a soft key indicates a desire for a new query. If so, the program returns to the connector "H" (see Fig. 4a). If not, it is determined in step 172 whether the soft key indicates a desire to exit. If the exit key is not selected, a pre-recorded message requesting that the caller select a new query or exit is played in step 174. If the exit key is selected, a pre-recorded session termination message is played and the program ends, releasing all the resources in step 176.

Fig. 4e illustrates the logic involved in providing a caller with another way of entering an index used as a database lookup when speech recognition fails. In step 178, the access server 16 sends soft key options to the ADSI display telephone 10 providing the caller with an opportunity to select another way of entering an index. The soft key options may be accompanied with a pre-recorded voice announcement instructing the caller respecting the options. The access server 16 monitors

- 24 -

the connection in step 180 to determine whether the caller has elected to spell the index verbally. If that option is selected, the access server 16 initializes for an alpha speech recognition algorithm in step 182 and
5 collects the voice signals in step 184 which are passed to the alpha speech recognition algorithm. Control is then passed to connector "I" (see Fig. 4b) and program execution proceeds as explained above with reference to Figs. 4b, 4c and 4d. If voice spelling is not selected
10 in step 180, it is determined whether manual spelling is selected in step 186. It will be understood by those skilled in the art that various incorrect key selection and time-out provisions are provided in the program but are not illustrated here for sake of simplicity.

15 If manual spelling is selected in step 186, the access server 16 collects the DTMF signals in step 188 and converts them to alpha characters as described above with reference to Fig. 3. Concurrently, the access server 16 monitors the connection for a soft key signal
20 indicating that the entire index has been spelled in step 190. If the soft key signal does not indicate that the spelling is complete, it is determined in step 192 whether the soft key is a request to exit. If so, the user is returned to connector "A" (see Fig. 4b) and
25 program execution continues. Otherwise, a predetermined time lapse is monitored in step 194. If the predetermined time is determined to have elapsed in step 194, a pre-recorded announcement is played in step 196 requesting that the caller indicate completion
30 of the manual spelling or indicate that the spelling is completed. When an indication of completion is

- 25 -

determined in step 190, the collected alpha characters are assembled into an index string in step 198 and the program returns to connector "C" (see Fig. 4c) where the index is stored in step 148 and the program execution
5 continues.

Fig. 5 shows an exemplary configuration for a 411 directory of residential numbers (see Figs. 1, 3) accessed by the apparatus in accordance with the invention. The 411 directory 26 is supported by at least
10 one computer having an operating system 200 which runs an Internet server application 202. A communications interface 204 permits the directory to be connected to the wide area network 24 (see Fig. 1). The Internet server receives queries through the communications
15 interface 204 from the wide area network 24. The queries are passed from the access server 16 in TCP/IP format, well known in the art. The Internet server 202 passes queries received from the access server 16 to an Application Programming Interface (API) 208 which
20 extracts one or more appropriate records from the database 210 in response to the query. A response message is formulated by the API 208 and returned to Internet server 202. The Internet server 202 addresses the response and sends it back to a querying access
25 server 16 (see Fig. 1), as was explained above with reference to Figs. 3 and 4a-e.

It will be understood by those skilled in the art that the methods and apparatus described above may be implemented in many applications and in many different
30 ways. Although the application described for purposes of illustration was the retrieval of information from a

- 26 -

database, and in particular, telephone numbers from a directory assistance service, this is only one of a great many applications for the methods and apparatus in accordance with the invention. For example, the
5 invention may be used to automate product sales, to make reservations for food or lodging, or to enable any other transaction which is usefully or preferably accomplished by voice over communications lines. The scope of the invention is therefore intended to be limited solely by
10 the scope of the appended claims.

- 27 -

CLAIMS:

1. A method of improving the utility of speech recognition of words spoken by a speaker in which a word spoken by the speaker is captured in electronic form and passed to a speech recognition algorithm (20) and at least one representation of the word is returned by the speech recognition algorithm C H A R A C T E R I Z E D in that:

the at least one representation of the word is displayed for the speaker as text to permit the speaker to select a correct representation of the word from among the at least one displayed representation.

2. A method as claimed in claim 1 wherein the words spoken by the speaker are converted to electronic form by a voice terminal.

3. A method as claimed in claim 2 wherein the voice terminal is a telephone (10).

4. A method as claimed in claim 3 wherein the telephone (10) is a display telephone which conforms to an Analog Display Services Interface (ADSI) standard.

5. A method as claimed in claim 1 wherein the electronic form of the word spoken by the speaker is captured by an access server (16) which is accessed by the speaker using a dial-up connection over a voice grade telephone line (12).

- 28 -

6. A method as claimed in claim 5 wherein the speech recognition algorithm (20) resides on another server connected to the access server (16) by a communications network (22).

7. A method as claimed in any preceding claim wherein if the speaker confirms that none of the at least one representations of the word displayed are representative of the word spoken, the speaker is prompted to verbally spell the word, and letters spoken by the speaker in response to the prompt are passed to an alpha speech recognition algorithm (20).

8. A method as claimed in claim 4 wherein if the speaker confirms that none of the at least one representations of the word displayed are representative of the word spoken, the speaker is prompted to manually spell the word using a dial pad of the display telephone (10).

9. A method as claimed in claim 8 wherein keys on the dial pad are pressed once for each relative position of a letter on a key of the dial pad in order to manually spell the spoken name, so that if the letter appears in a first position on the key, the key is pressed once to represent the letter, if the letter is in a second position on the key, the key is pressed twice to represent the letter and if the letter is in a third position on the key, the key is pressed three times to represent the letter.

- 29 -

10. A method as claimed in any preceding claim wherein the method is used for automating telephone directory services for a telephone user having a display telephone (10), wherein the telephone user is prompted for names used as indicia to locate an entity in a telephone directory (26-34), and a spoken name is accepted for each index and passed to a speech recognition algorithm (20) that returns at least one representation of the spoken name, each representation being displayed as text on the display telephone (10) to permit the speaker to select a correct representation of the spoken name; and, a query to the telephone directory is assembled after a correct representation of each index has been selected by the speaker in order to retrieve a record for the entity from the telephone directory.

11. A method as claimed in claim 10 further comprising a step of providing the speaker with another way of providing an index in an event that the index cannot be recognized by the speech recognition algorithm (20).

12. A method as claimed in claim 11 wherein the other way of providing the index comprises enabling the speaker to verbally spell the spoken name.

13. A method as claimed 11 wherein the other way of providing the index comprises enabling the speaker to manually spell the name using a dial pad of the display telephone (10).

- 30 -

14. A method as claimed in any one of claims 10-13 wherein the steps of prompting, accepting, passing and assembling are accomplished by an access server (16) that is accessed by the telephone user by dialing a predetermined telephone number.

15. A method as claimed in claim 14 wherein the step of displaying is accomplished by passing the representations from the access server (16) to the display telephone (10) over a telephone line (12) along with commands that enable the display telephone to display the representations as text for the speaker.

16. Apparatus for improving the utility of speech recognition of words spoken by a speaker, comprising a computer (16) enabled to receive voice and data signals over a communications link (18), the computer being programmed to prompt the speaker for spoken words which are received from the communications link as voice signals and to pass the spoken words to a speech recognition algorithm (20) which returns representations of the spoken words to the computer CHARACTERIZED in that:

the computer is further enabled to pass the representations of the spoken words to a voice terminal with a display surface which displays the representations for the speaker to permit the speaker to select a correct representation of the spoken words.

17. Apparatus as claimed in claim 16 wherein the computer is a server (16) in a local area network (22)

- 31 -

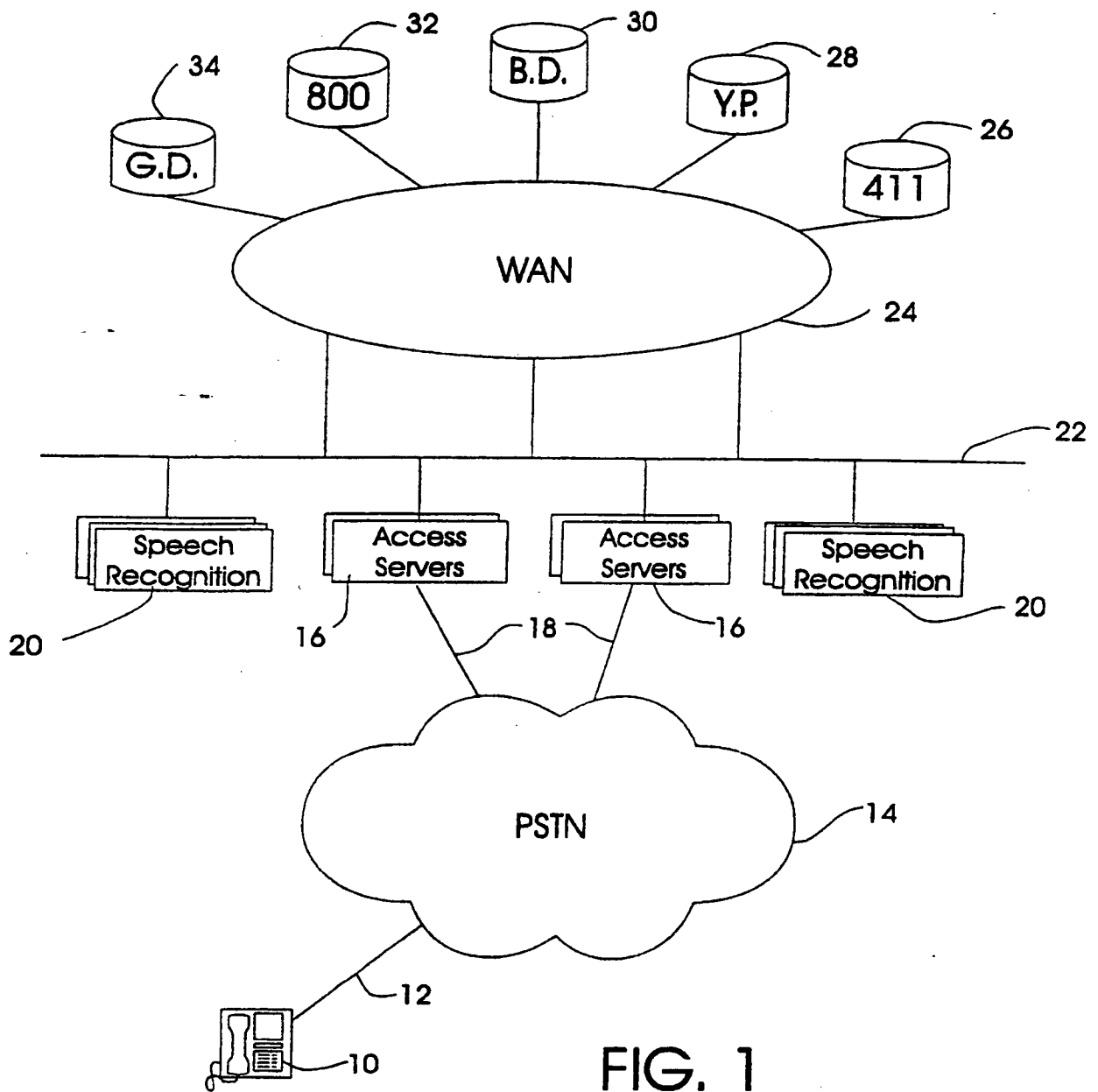
and the server is connected to a switched telephone network (14) by a voice grade connection (18).

18. Apparatus as claimed in claim 17 wherein the speech recognition algorithm (20) resides on another server connected to the local area network (22).

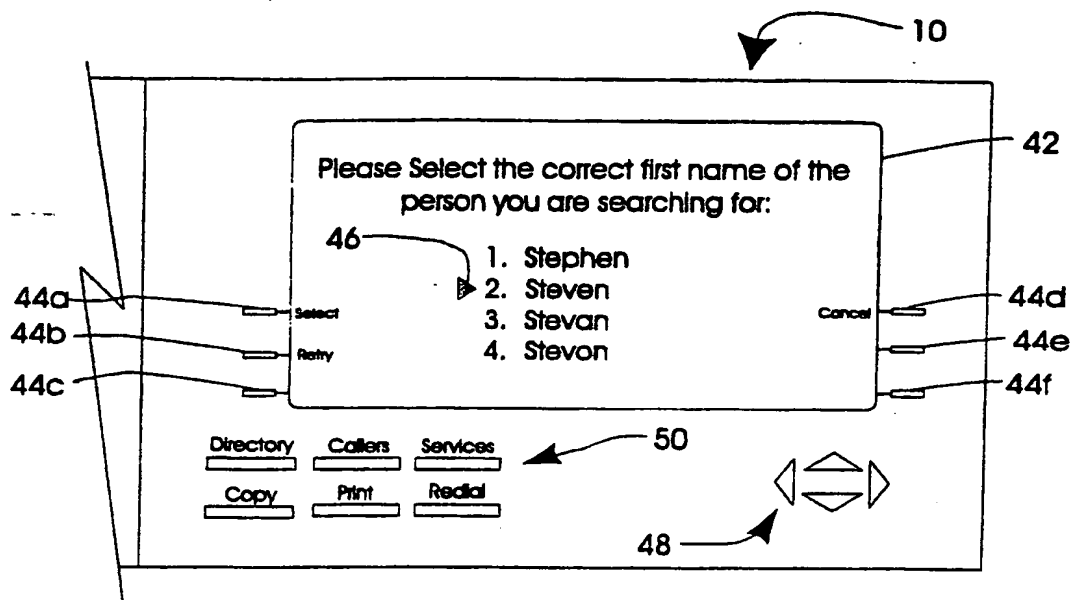
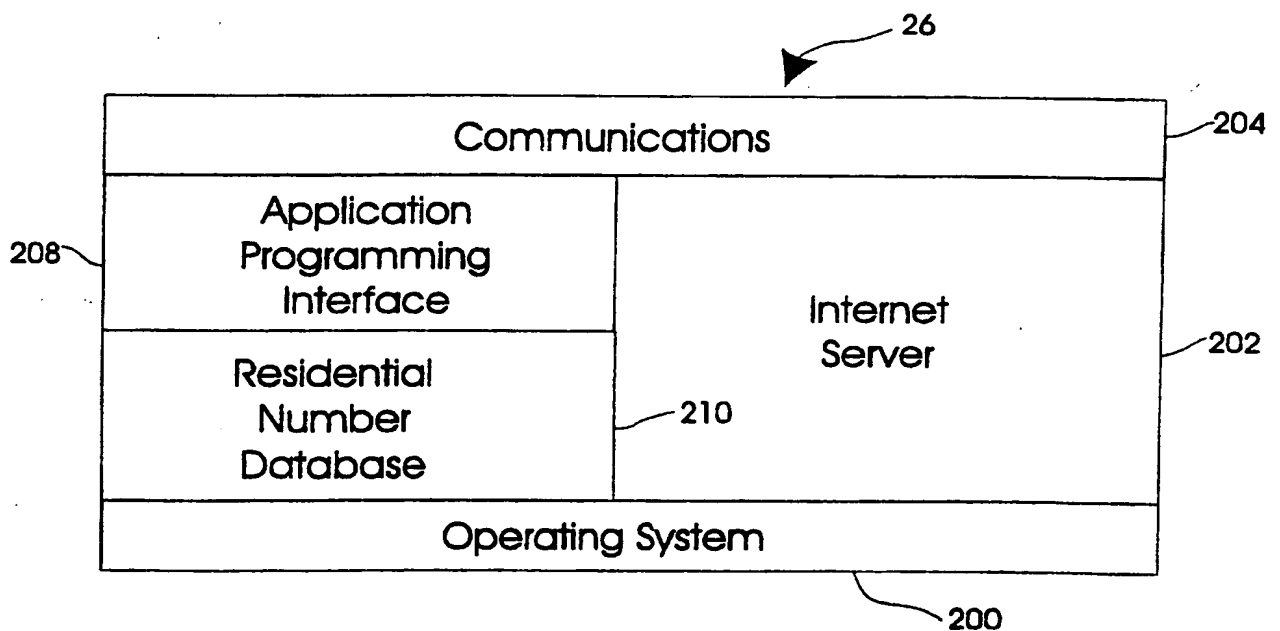
19. Apparatus as claimed in claims 17 or 18 wherein the apparatus is used to provide automated telephone directory (26-34) services and the spoken words are used as indicia for retrieving subscriber information from a telephone directory connected to a wide area network (24) which may be accessed by the access server (16).

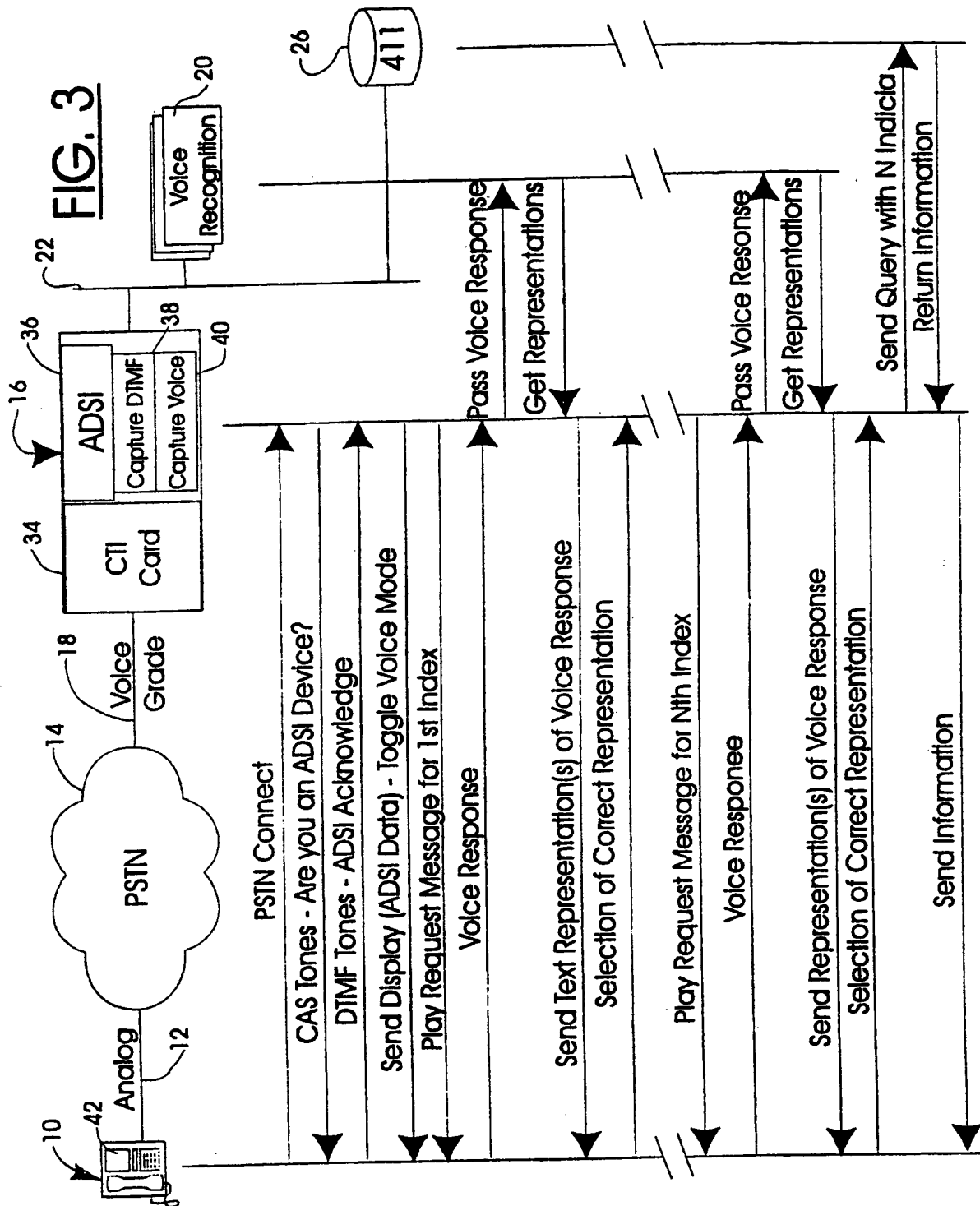
20. Apparatus as claimed in any one of claims 17-19 wherein the access server (16) may selectively pass words to an alpha speech recognition algorithm to enable the speaker to verbally spell a spoken name if a spoken version of the spoken name cannot be interpreted by the speech recognition algorithm (20).

1/8



2/8

FIG. 2FIG. 5



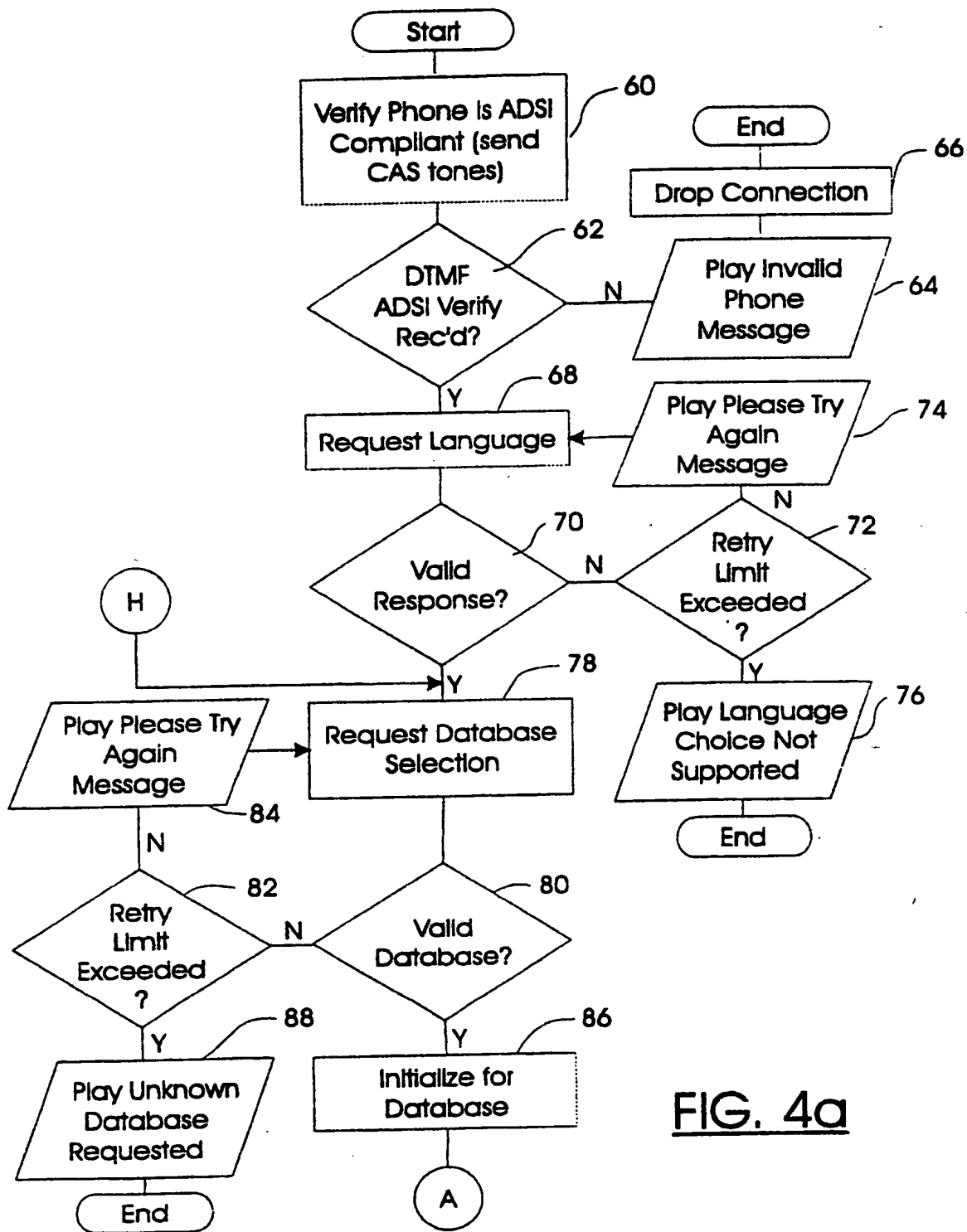


FIG. 4a

5/8

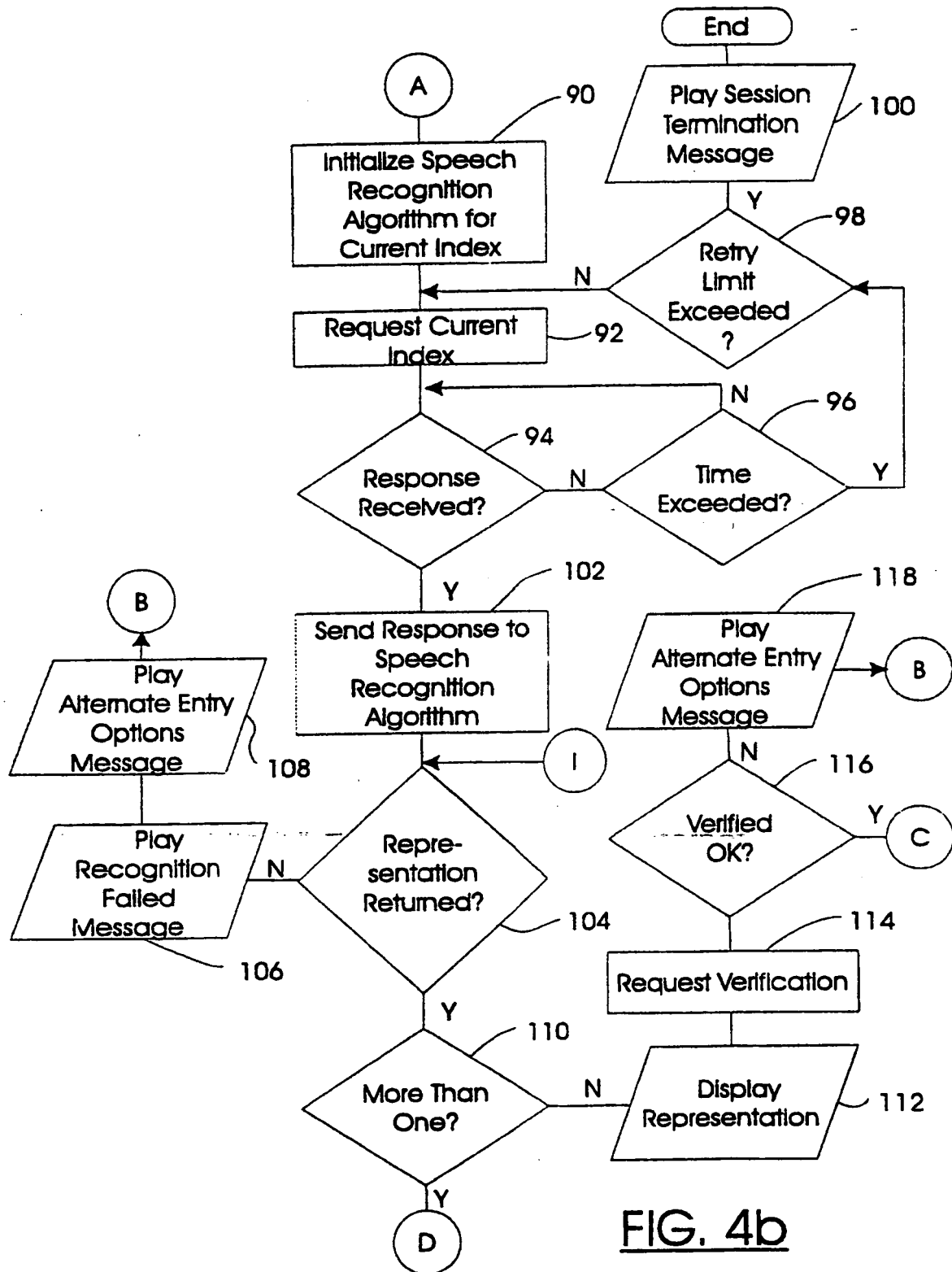


FIG. 4b

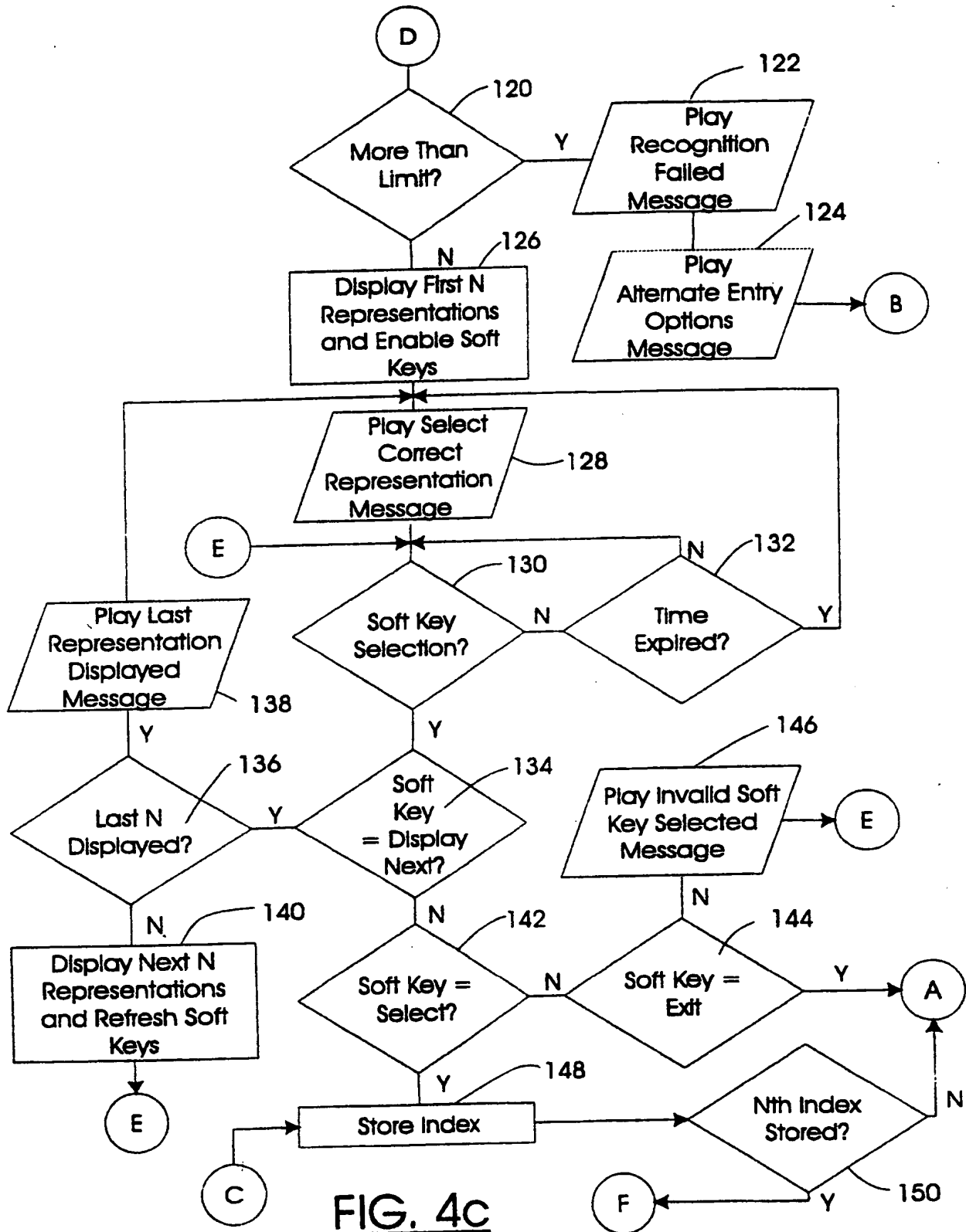
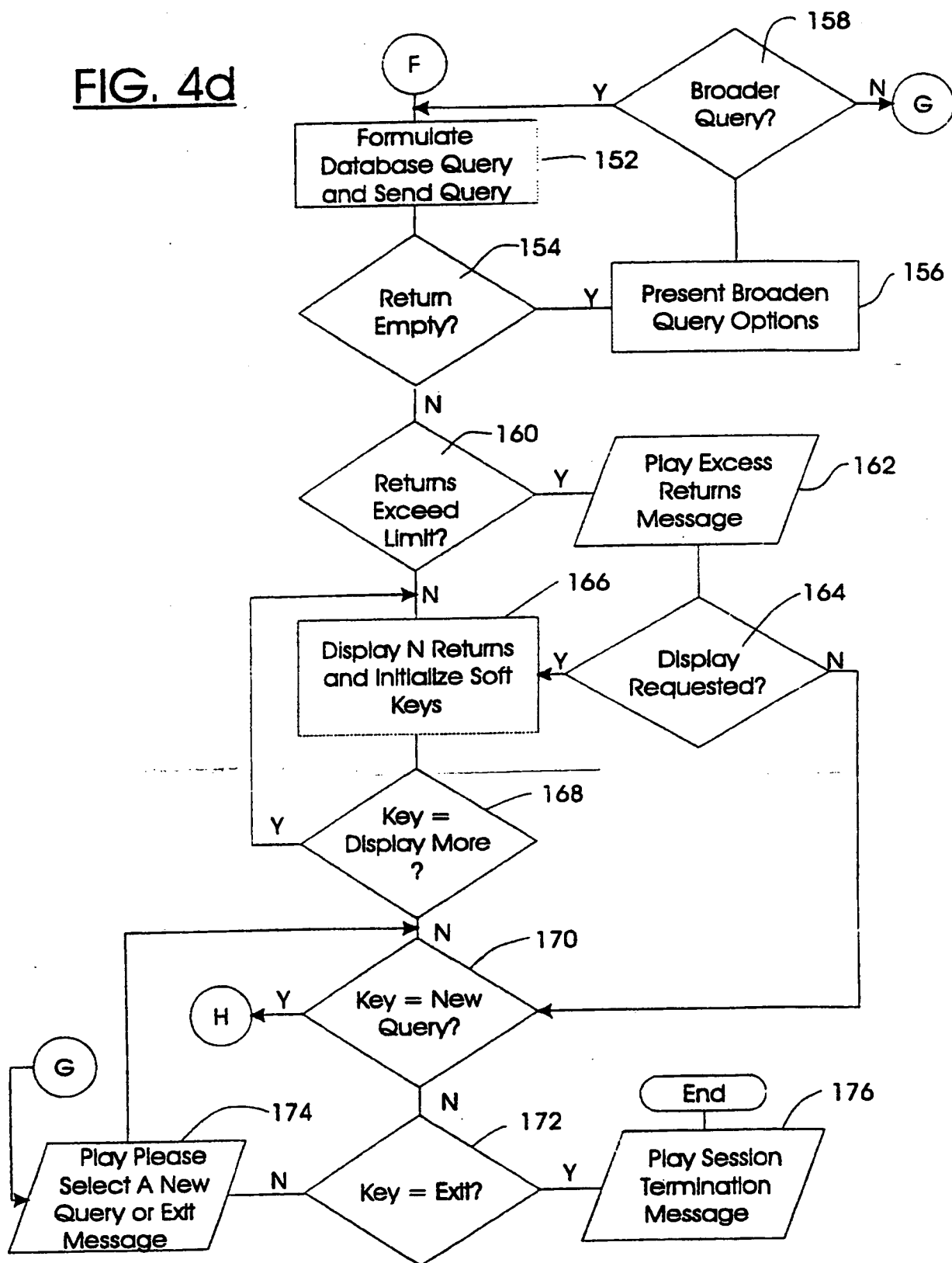
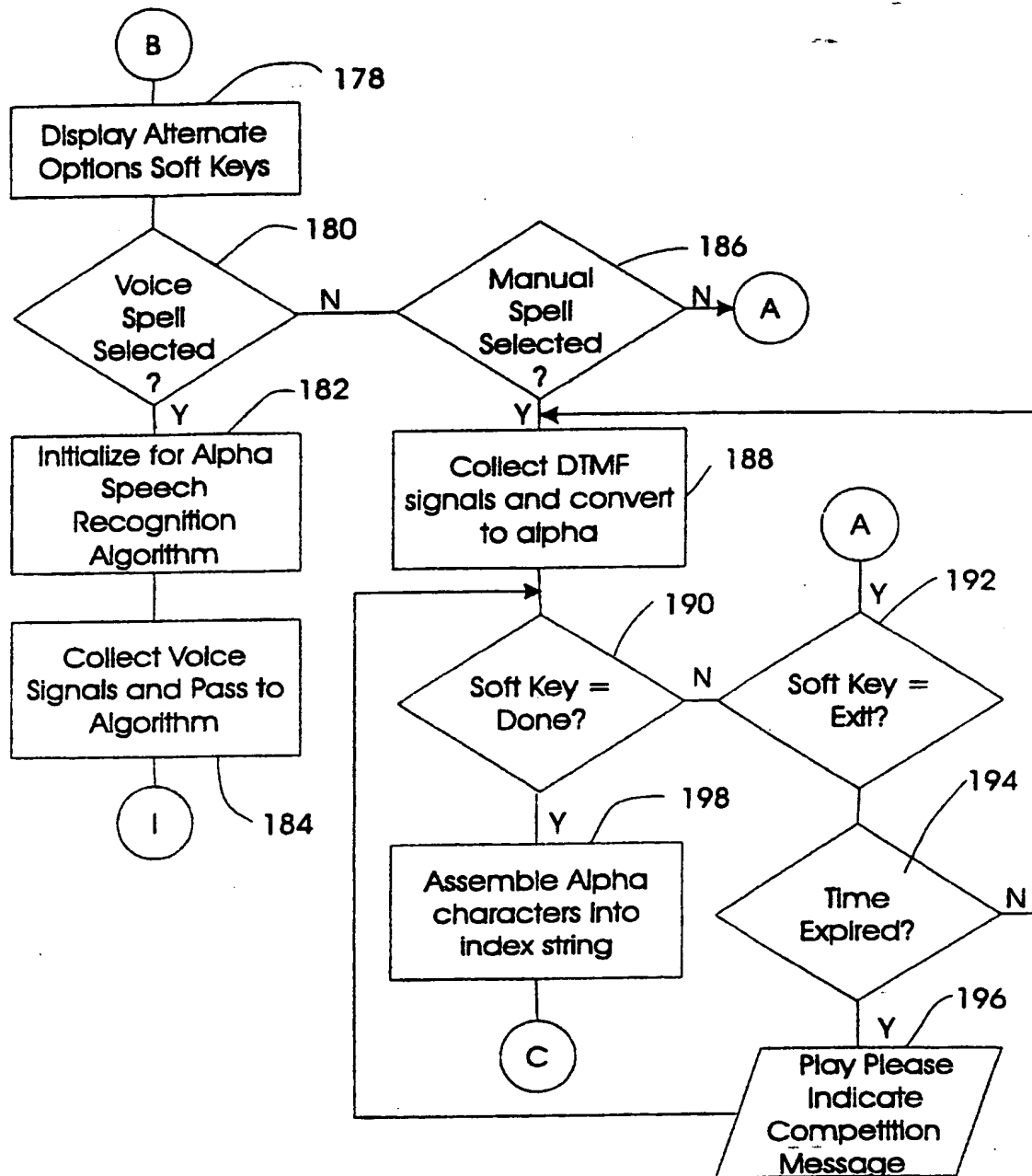


FIG. 4c

FIG. 4d

FIG. 4e

INTERNATIONAL SEARCH REPORT

International Application No
PCT/CA 98/00978

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G10L5/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G10L H04M G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	"METHOD FOR LOWERING BRANCHING FACTORS IN FORMS FILLING APPLICATIONS" IBM TECHNICAL DISCLOSURE BULLETIN, vol. 38, no. 7, July 1995, page 525/526 XP000521777 see the whole document	1, 7, 16, 17
A, P	EP 0 823 809 A (AT & T CORP) 11 February 1998 see abstract; claims 1, 2, 4, 15; figure 1	2-6, 17, 18
A	US 3 928 724 A (BYRAM HAROLD E ET AL) 23 December 1975 see abstract; claim 20; figure 1	2-6, 17, 18
	-/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

8 December 1998

Date of mailing of the international search report

18/12/1998

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Van Doremalen, J

INTERNATIONAL SEARCH REPORT

In tional Application No
PCT/CA 98/00978

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>WO 94 14720 A (HUBER GES FUER VERSORGUNGS UND ;HUBER STEPHAN (DE); HUBER ERICH (D) 7 July 1994 see abstract</p> <p>-----</p>	<p>7,8,12, 13,20</p>

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/CA 98/00978

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 0823809	A	11-02-1998	CA 2201067 A JP 10093622 A	06-02-1998 10-04-1998
US 3928724	A	23-12-1975	NONE	
WO 9414720	A	07-07-1994	DE 4244249 C EP 0675859 A	01-06-1994 11-10-1995